SNP_TATA_Comparator: genomewide landmarks for preventive personalized medicine

Mikhail Ponomarenko¹,², Dmitry Rasskazov¹, Irina Chadaeva¹,², Ekaterina Sharypova¹, Petr Ponomarenko³, Olga Arkova¹,², Elena Kashina¹, Nikita Ivanisenko¹, Dmitry Zhechev¹, Ludmila Savinkova¹, Nikolay Kolchanov¹,²

¹Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk 630090 Russia; ²Department of Natural Sciences, Novosibirsk State University, Novosibirsk 630090, Russia; ³Children’s Hospital Los Angeles, University of Southern California, CA 90027, USA; ⁴Vector-Best Inc., Koltsovo, Novosibirsk Region 630559, Russia

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. WEB service SNP_TATA_Comparator
   3.1. How to use WEB service SNP_TATA_Comparator
   3.2. Bioinformatics model
   3.3. Hardware platform and software design
4. Examples of known and candidate SNP markers (of hereditary diseases) that can alter TATA boxes of human genes
   4.1. The human MBL2 gene
   4.2. Human HBB and HBD genes
   4.3. The human IL1B gene
   4.4. The human NOS2 gene
   4.5. The human F3 gene
   4.6. The human LEP gene
   4.7. The human ABCA9 gene
5. Preclinical verification in vitro and ex vivo of the candidate SNP markers of hereditary diseases
   5.1. Verification of candidate biomedical SNP markers using an electrophoretic mobility shift assay (EMSA) in vitro
   5.2. Verification of candidate biomedical SNP markers in real time in vitro using high-resolution spectrometer SX.20
   5.3. Verification of candidate biomedical SNP markers using human cell culture under ex vivo conditions
6. Biomedical predictions made using a Key Words search for known biomarkers of a human disease in databases
   6.1. A Key Words search for known physiological markers of a human disease in the PubMed database
   6.2. Examples of SNP markers of hereditary diseases as candidate SNP markers of sporadic Alzheimer’s disease (AD)
   6.3. Examples of SNP markers of hereditary diseases as candidate SNP markers of autoimmunity-related comorbidities
   6.4. Examples of SNP markers of hereditary diseases as candidate SNP markers of circadian-rhythm-related comorbidities
   6.5. Examples of SNP markers of hereditary diseases as candidate SNP markers of obesity-related comorbidities
   6.6. Examples of SNP markers of hereditary diseases as candidate SNP markers of aggressiveness-related complications
7. Summary and perspective
8. Acknowledgments
9. References
1. ABSTRACT

Year after year, conditions, quality, and duration of human lives have been improving due to the progress of science, technology, education, and medicine, which however has a downside. Owing to improvement in children’s nutrition, developmental acceleration occurs that imbalances a child’s system. Because of virtual worlds of the Internet, social experience of teenagers expands and clashes with puberty of adolescents. Due to the comfort of cities, urbanization emerges and causes stress to adults because of artificial light, noise, pollution, violations of personal space, and family disruption. At old age, all these factors taken together contribute to loneliness, cancer, diabetes, drug addiction, and sporadic Alzheimer’s disease, which shorten the lifespan, as reviewed in the US, 1990–2010 (1). That is why, a person may ask oneself: “What can I do now to keep my health in my old age?” To help them, we provide this comprehensive review on predictive preventive personalized medicine. This branch of molecular medicine uses single nucleotide polymorphisms to prevent diseases on the basis of the difference between the individual and reference human genomes.

2. INTRODUCTION

Year after year, conditions, quality, and duration of human lives have been improving due to the progress of science, technology, education, and medicine, which however has a downside. Owing to improvement in children’s nutrition, developmental acceleration occurs that imbalances a child’s system. Because of virtual worlds of the Internet, social experience of teenagers expands and clashes with puberty of adolescents. Due to the comfort of cities, urbanization emerges and causes stress to adults because of artificial light, noise, pollution, violations of personal space, and family disruption. At old age, all these factors taken together contribute to loneliness, cancer, diabetes, drug addiction, and sporadic Alzheimer’s disease, which shorten the lifespan, as reviewed in the US, 1990–2010 (1). That is why, a person may ask oneself: “What can I do now to keep my health in my old age?” This personal initiative in health maintenance is one of the main reasons behind predictive preventive personalized medicine (2). This newest branch of molecular medicine uses single nucleotide polymorphisms to prevent diseases on the basis of the difference between the individual and reference human genomes.

Before the genomics era, discovery of the associations between SNPs and diseases was mostly due to luck (3, 4). Now, in the postgenomic era, such studies are systematic and genomewide within the framework of the large worldwide project “1000 Genomes” (5). The results of the project are stored in the dbSNP database as true SNPs (6). This database is an inherent part of the reference human genome and contains ancestral alleles of all SNPs, whereas the human variome represents their minor alleles. Currently the database contains over 8.98 billion known unannotated SNPs (7), which can be mapped using the UCSC Genome Browser (8). In turn, biomedical databases GWAS (genomewide association study) (9), OMIM (10), ClinVar (11), and HapMap (12) annotate these SNPs by documenting associations with hereditary diseases and with their complications and comorbidities (e.g., (13)). Each such association is characterized by the statistical significance of the discrimination between a cohort of patients with a given disease and healthy volunteers (as a control), with the SNPs serving as biomarkers of this disease according to a conventional clinical protocol (14). The vast majority of SNPs are neutral while only a tiny minority is biomedical SNP markers as was postulated by the theory of neutral evolution (15) and Haldane’s dilemma (16). Therefore, the clinical search for SNP markers of diseases by “trial and error” is very slow, costly, and labor-intensive.

Most of currently known biomedical SNP markers have been found within protein-coding gene regions where their manifestations are invariant and easily detectable because of the disruption in both structure and function of the altered protein (10). Nevertheless, neither medication nor a lifestyle change can fully correct pathological effects of these SNPs; that is why they are rather valuable for physicians but not for people who would like to prevent various diseases by sequencing their individual genome. In contrast, the smallest number of biomedical markers has been found in the regulatory regions of genes, where the mutations alter only the concentration of a protein in human cells without any damage to the structure and function of this protein (17). Because nothing changed in proteins but their concentrations, the pathological effects of these regulatory biomedical SNP markers should be correctable by medications and/or lifestyle changes. For this reason, such biomedical SNP markers can be interesting both to physicians (may help to improve patient care) and to people who would like to prevent such diseases by sequencing their individual genome.

The majority of the known regulatory biomedical SNP markers alter the binding site for TATA-binding protein (TBP) because they are located within the small region (−70; −20) upstream of the transcription start site (18). These SNP markers are readily detectable due to the positive correlation between the expression level of the human gene containing them and the affinity of TBP for the promoter of this gene (19). This is because TBP’s binding to its binding site is the very first obligatory molecular event of transcription initiation in eukaryotes (20, 21).
For the remaining ~2,600 DNA-binding proteins (22), experimental data on their binding to human gene promoters in health (23), in disease (24), and after treatment (25) are still being accumulated.

Computer-based elimination of neutral SNPs among the known unannotated SNPs (7) may accelerate and facilitate the clinical search for biomedical SNP markers (4). This task can be performed by means of many publicly available Web services performing well only on some SNPs and diseases ((26–42); for a review, see (43, 44)). The most advanced software appears to be targeted to the dynamics of molecular structures and allows researchers to estimate the SNP-caused changes in proteins (e.g. (45)). The new trend in this active field of research is Web navigation services that help users to generate their own ideas on how an SNP of interest can affect symptoms of diseases under study by predicting how this SNP can alter the protein’s structure-function relation (46). Another innovation is Web service PredictSNP2 (47). It expresses the numerical estimates of alteration in a protein’s structure-function relations caused by an SNP as qualitative categories of human health (47). These categories are more understandable to the general population.

Meanwhile, the smallest progress has been made on regulatory SNPs because their manifestations may vary from cell to cell, from tissue to tissue, from patient to patient, and from subpopulation to subpopulation without any changes in a protein’s structure-function relations (48). That is why development of the approaches to computer-based prediction of candidate SNP markers of human hereditary diseases within regulatory gene regions is a challenging problem that limits the progress of predictive preventive personalized medicine (2).

Trying to find an answer to the question “What can be done to keep one’s health at old age?” here we provide a comprehensive review on the topic of known and candidate regulatory SNP markers of hereditary diseases. This review is focused on the markers near the TBP-binding sites of human gene promoters because these regulatory sites have been studied the most. For example, one of such candidate SNP markers is rs72661131, which is associated with stroke and sporadic AD (49). Risks of both diseases can be minimized by adding some natural marine products into the diet of an individual who has a minor allele of this SNP (50). Of course, this is an oversimplification: if stroke and sporadic AD could be prevented by a natural marine diet alone, the life of the general population would be easy. Nevertheless, this candidate SNP marker may be interesting for people who would like to bring their lifestyle in line with their sequenced individual genome. Thus, we focused this review on a public Web service called SNP_TATA_Comparator (http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl), which was developed specifically for predicting the candidate SNP markers near the TBP-binding sites of human gene promoters (51).

3. WEB-SERVICE SNP_TATA_COMPARATOR

The Web-service SNP_TATA_Comparator, http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl, calculates a computer-based statistical estimation of SNP-caused alteration of TBP’s binding affinity for a human gene promoter. Based on this estimate, the Web-service SNP_TATA_Comparator predicts a change in the expression of the genes that can be associated with diseases, their comorbidities, and complications (51). Let’s see how to use it and review its bioinformatics model, hardware platform, and software design in more detail.

3.1. How to use WEB service SNP_TATA_Comparator

Figure 1 illustrates how a person can use SNP_TATA_Comparator (51) in practice. As a preliminary preparation, he or she should first find the human gene, the promoter of this gene, and the SNPs of this promoter, using another public Web service “UCSC Genome Browser” (8) as shown in Figure 1 A.

Next, using the BioPerl library (52), the user can extract the promoter sequences containing these SNPs from the reference human genome (5) by means of Web service SNP_TATA_Comparator (51) as shown in Figure 1 B. Here, the “Base sequence” textbox contains the ancestral allele of the promoter of interest. Another “Editable sequence” textbox contains a copy of this DNA sequence, where the user should manually create the minor allele of interest according to its description from the dbSNP database (6) available via Web service “UCSC Genome Browser” (8). Clicking on the “Calculate” button transforms these two DNA sequences as the input data into the output data appearing in the “Result” textbox as shown in Figure 1, B and C. This transformation is based on a bioinformatics model shown in Figure 2 and described in detail in section 3.2. Finally, the line “DECISION” shows the prediction made by SNP_TATA_Comparator (51) using this model. Additionally, the line “Z-score” shows the p value of the probability rate of this prediction (where: alpha = 1 - p is its statistical significance).

3.2. Bioinformatics model

The bioinformatics model of the Web service SNP_TATA_Comparator (51) takes into account three steps of TBP’s binding to a eukaryotic gene promoter, namely: (i) TBP slides along DNA (53–55) <=> (ii) TBP stops at a TBP-binding site (56; 57) <=> (iii) the TBP–promoter complex is fixed by DNA helix’s bending to
Genomic landmarks for preventive personalized medicine

The 90° angle (58–60). This binding affinity can be estimated using the following empirical equation:

\[-\ln(K_0) = 10.9 - 0.2 \cdot (\ln(K_{\text{SLIDE}}) + \ln(K_{\text{STOP}}) + \ln(K_{\text{BEND}})),\]  

where 10.9 (in units) is nonspecific TBP–DNA affinity \(10^{-6}\) M (61), 0.2 is the stoichiometric coefficient (62), and \(K_{\text{STOP}}\) is the maximal score-value of Bucher’s position-weight matrix, which is the commonly accepted criterion of the TATA box: the canonical form of the TBP-binding site (57).

In Eq. (1), \(K_{\text{SLIDE}}\) is our empirical estimate of the equilibrium constant of TBP’s sliding along DNA (this constant was determined experimentally (53)), namely:

\[-\ln(K_{\text{SLIDE}}) = \text{MEAN}_{10bp} \{0.8(TA)_{3}\text{HALF} - 3.4\text{MinorGrooveWidth}_{\text{CENTER}} - 35.1\},\]

Figure 1. Known and candidate SNP markers of variable immunodeficiency (77), stroke (78), preeclampsia (79), and (likely) cardiovascular risk in rheumatoid arthritis (80) near TBP-binding sites of the human MBL2 gene promoter. (A) Unannotated SNPs (analyzed in this study) in the region (–70; –20) (where all proven TBP-binding sites (boxed) are located; double-headed arrow) of the human MBL2 gene promoter. These data were retrieved from dbSNP (6) using the UCSC Genome Browser (8). Dash-and-double-dot arrows: known and candidate biomedical SNP markers are predicted by a significant change in the affinity of TBP for the human MBL2 gene promoter. (B, C) The results produced by Web service SNP_TATA_Comparator (51) for the two SNP markers of the hereditary diseases in question: known rs72661131 (77–79) and candidate marker rs567653539 near the known TATA box (boxed) of the human MBL2 gene promoter. Solid, dotted, and dashed arrows indicate queries for the gene list, list of transcripts of a given gene, and DNA sequence of the promoter corresponding to the specified transcript in terms of the BioPerl library (52) of the reference human genome (5), respectively. Dash-and-dot arrows: estimates of significance of the alteration of gene product abundance in patients with the minor allele (mt) relative to the norm (ancestral allele, wt) expressed as a Z-score (the ratio of the difference between two means to the square root of the sum of the squares of their standard errors) using package R (66). Circles indicate the ancestral (wt) and minor (mt) alleles of the SNP marker labeled with its dbSNP ID (6).
Genomic landmarks for preventive personalized medicine

where (TA)$_{3'\text{HALF}}$ is the total amount of dinucleotide TA within the 3' half of the DNA sequence being analyzed; MinorGrooveWidth$_{\text{REGION}}$ is the mean width of the minor groove of the B-form of the DNA helix (54); 0.8, -3.4, and -35.1 are linear regression coefficients (62); MEAN$_{\text{TATA-box}}$ is the mean arithmetic value for all possible positions and orientations of the TBP-binding site, 15 bp long (this value was determined empirically (53)).

In Eq. (1), $K_{\text{BEND}}$ is our empirical estimate of the equilibrium constant at the DNA helix bending step on the basis of the macromolecular dynamics computations (58) describing how TBP can bind to DNA, namely:

$$-\ln(K_{\text{BEND}}) = \text{MEAN}_{\text{TATA-box}} \{0.9\text{WR}_{\text{FLANK}} + 2.5\text{TV}_{\text{CENTER}} + 14.4\}, \tag{3}$$

where WR = (TA, AA, TG, AG), TV = (TA, TC, TG) (63) (the IUPAC-IUB nomenclature (64)); 0.9, 2.5, and 14.4 are linear regression coefficients taken from ref. (63); MEAN$_{\text{TATA-box}}$ is the mean arithmetic value for both DNA strands of the TBP-binding site at the position of the maximal score-value of Bucher’s position-weight matrix (57).

Additionally, the estimates of standard deviation of the $-\ln(K_D)$ (Eq. 1)—for 78 possible substitutions, $s_{i,j}$→X, at each j-th position (-14 < j < 13; 78 = 3 * 26) within the 26-bp DNA window—was heuristically estimated as

$$d = \left(\frac{\sum_1^{j-1} \sum_1^{10} (x_{j}-\bar{x})^2}{\sum_1^{j-1} \sum_1^{10} (x_{j}-\bar{x})^2} \right)^{1/2}, \tag{4}$$

This equation (4) estimates the resistance of the TBP-binding site of the promoters against the majority of single-nucleotide substitutions (65).

Next, the results of calculations according to Eqs. (1–4) for the promoter DNA sequences of both ancestral (wt) and minor (mt) alleles of a given gene, $-\ln(K_{D,\text{wt}})$ and $-\ln(K_{D,\text{mt}})$, respectively, are compared with each other in terms of Fisher’s Z-score as described in ref. (66):

$$Z = \frac{|\ln(K_{D,\text{wt}})/K_{D,\text{mt}}|}{(d_{\text{wt}}^2 + d_{\text{mt}}^2)^{1/2}}. \tag{5}$$

Then, the Z-value obtained is converted using R statistical package (66) into a p value, i.e., the probability rate (where alpha = 1 - p is the statistical significance) as shown in Figure 1 B.

Finally, the following decision is made:

IF $-\ln(K_{D,\text{mt}})$ is statistically significantly greater than $-\ln(K_{D,\text{wt}})$,

THEN {DECISION is “there is overexpression of the minor allele of this gene in comparison with the ancestral allele”};

ELSE (IF $-\ln(K_{D,\text{mt}})$ is statistically significantly less than $-\ln(K_{D,\text{wt}})$),

THEN {DECISION is “there is underexpression of the minor allele of this gene in comparison with the ancestral allele”},

OTHERWISE {DECISION is “alteration of the expression of this gene is insignificant”} \tag{6}

For each SNP thus analyzed, the decision (Eq. 6) is the main result shown in the line “DECISION” of the “Result” textbox, and the p value can be found in the line “Z-score” (Figures 1, B and C).

It should be noted that before developing software Web-service SNP_TATA_Comparator (51) based on equations (1–6), the three-step molecular mechanism for the binding between TBP and the TATA box was predicted in silico (67) and detected in vitro.
Genomic landmarks for preventive personalized medicine

SNP

3.3. Hardware platform and software design

Web service SNP_TATA_Comparator (51) is a bioinformatics application installed on the hardware HP 380DL G6 under both virtualization “Citrix XenServer” and operating “CentOS 6” systems, which connect the Web server “Ngix + Apache,” which is supported by the Siberian Supercomputer Center (Novosibirsk, Russia). It takes queried data from the reference human genome (5) using the BioPerl toolkit (52) as shown in Figure 1 B. Its bioinformatics model is the executable applet encoded primarily in the programming language C of the ANSI standard, which uses standard statistical software R (66). The user interface is written using standard Java-script. This Web service is publicly available (http://beehive.bionet.nsc.ru/cgi-bin/mgs/tatascan/start.pl).

4. EXAMPLES OF KNOWN AND CANDIDATE SNP MARKERS (OF HEREDITARY DISEASES) THAT CAN ALTER TATA-BOXES OF HUMAN GENES

Table 1 shows the selected examples of known and candidate SNP markers of hereditary diseases. These SNPs can alter TATA-boxes in the canonical form of the TBP-binding sites of human gene promoters as predicted by the Web service (51). Here, we consider only one example from these illustrative results in detail, and the other SNPs are presented briefly.

4.1. The human MBL2 gene

The human MBL2 gene (soluble mannose-binding lectin 2) contains one known SNP marker (rs72661131) of variable immunodeficiency (77), stroke (78), and preeclampsia (79). This SNP is a substitution of a minor c for an ancestral T at position –39 (hereafter denoted as –39T>c) in the promoter of this gene as shown in Table 1. Near this known SNP marker, we found unannotated rs549858786, which contains the most widely studied SNP marker, we found unannotated rs549858786, which can cause a hemoglobin deficiency. Thus, they can be considered candidate SNP markers of the same pathologies (Table 1).

4.2. Human HBB and HBD genes

Human HBB and HBD genes (beta- and delta-chains of hemoglobin, respectively) contain the largest number of SNP markers (rs34500389, rs33981098, rs33980857, rs34598529, rs33931746, rs397509430, and rs35518301) of resistance to malaria and thalassemia caused by underexpression of these genes (81). Near these biomedical SNP markers, there are three unannotated SNPs (rs63750953, rs281864525, and rs34166473) that can cause a hemoglobin deficiency. Thus, they can be considered candidate SNP markers of the same pathologies (Table 1).

4.3. The human IL1B gene

The human IL1B gene (interleukin 1 beta) contains the most widely studied SNP marker (rs1143627) of gastric ulcer, chronic gastritis, gastric cancer, hepatocellular carcinoma, non–small cell lung cancer, Graves’ disease, obesity, and major depressive disorder (82–88), which were clinically shown to be associated with overexpression of this gene as shown in Table 1. Near this known SNP marker, we found unannotated rs549858786, which was reported to lower expression (Table 1), whereas an IL1B protein deficiency is a biochemical marker of rheumatoid arthritis (89), whose candidate SNP marker can thus be rs549858786.

4.4. The human NOS2 gene

The human NOS2 gene (inducible nitric oxide synthase 2) contains an SNP marker of resistance to malaria as a beneficial effect (90) and epilepsy as an adverse effect (91). It is a –51T>C substitution (position 27800550 of the complementary DNA strand on chromosome 17 relative to the transcription start site of this gene (90)) that causes NOS2 overexpression according to clinical research (90, 91).
Table 1. Known SNP markers altering affinity of the TATA-binding protein (TBP) for human gene promoters; their SNP neighbors

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP (6) or GRCh38/hg38 (Reference)</th>
<th>5'-flank</th>
<th>wt$^2$</th>
<th>3'-flank</th>
<th>$K_{\text{wt}}$/$K_{\text{mt}}$</th>
<th>SNP-caused change$^4$</th>
<th>$\kappa_{\text{SNP}}$</th>
<th>Z$^5$</th>
<th>A$^6$</th>
<th>Known diseases (known SNP markers) or predicted diseases (candidate SNP markers; Figure 7: Result-1)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBL2</td>
<td>rs72661131 tctatttct tc</td>
<td></td>
<td></td>
<td></td>
<td>1.74 4.19</td>
<td>Down</td>
<td>-0.87</td>
<td>12</td>
<td>10$^6$</td>
<td>variable immunodeficiency, stroke, preclampsia, and (likely) cardiovascular risk in rheumatoid arthritis (77–79), (51, 80)</td>
<td></td>
</tr>
<tr>
<td>MBL2</td>
<td>rs562962093 aatattttct ag</td>
<td></td>
<td></td>
<td></td>
<td>1.74 5.28</td>
<td>Down</td>
<td>-1.11</td>
<td>15</td>
<td>10$^6$</td>
<td>(likely) variable immunodeficiency, stroke, cardiovascular risk in rheumatoid arthritis (51, 77–79, 80)</td>
<td></td>
</tr>
<tr>
<td>MBL2</td>
<td>rs567653539 ttcttataa ga</td>
<td></td>
<td></td>
<td></td>
<td>1.74 1.25</td>
<td>Up</td>
<td>0.33</td>
<td>12</td>
<td>10$^6$</td>
<td>(likely) cardiovascular risk in rheumatoid arthritis (51, 80)</td>
<td></td>
</tr>
<tr>
<td>HBB</td>
<td>rs33980857 gggctgggca a.t.c</td>
<td></td>
<td></td>
<td></td>
<td>4.59 21.18</td>
<td>Down</td>
<td>-1.51</td>
<td>27</td>
<td>10$^6$</td>
<td>malaria resistance, thalassemia (81)</td>
<td></td>
</tr>
<tr>
<td>HBB</td>
<td>rs34598529 ggcctgggca a.g.c</td>
<td></td>
<td></td>
<td></td>
<td>4.59 17.69</td>
<td>Down</td>
<td>-1.35</td>
<td>24</td>
<td>10$^6$</td>
<td>malaria resistance, thalassemia (81)</td>
<td></td>
</tr>
<tr>
<td>HBB</td>
<td>rs33931746 gcgtgggaca a.g.c</td>
<td></td>
<td></td>
<td></td>
<td>4.59 10.84</td>
<td>Down</td>
<td>-0.87</td>
<td>14</td>
<td>10$^6$</td>
<td>malaria resistance, thalassemia (81)</td>
<td></td>
</tr>
<tr>
<td>HBB</td>
<td>rs63750953 ctgggcaata aa.g.c</td>
<td></td>
<td></td>
<td></td>
<td>4.59 8.28</td>
<td>Down</td>
<td>-0.60</td>
<td>9</td>
<td>10$^6$</td>
<td>(likely) malaria resistance, thalassemia (51, 81)</td>
<td></td>
</tr>
<tr>
<td>HBB</td>
<td>rs281864525 tgggcaataa ag.c</td>
<td></td>
<td></td>
<td></td>
<td>4.59 7.34</td>
<td>Down</td>
<td>-0.46</td>
<td>7</td>
<td>10$^6$</td>
<td>(likely) malaria resistance, thalassemia (51, 81)</td>
<td></td>
</tr>
<tr>
<td>HBD</td>
<td>rs35518301 caggacacgc a.c</td>
<td></td>
<td></td>
<td></td>
<td>4.19 7.95</td>
<td>Down</td>
<td>-0.63</td>
<td>11</td>
<td>10$^6$</td>
<td>malaria resistance, thalassemia (81)</td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td>rs34166473 aggacacgca tc</td>
<td></td>
<td></td>
<td></td>
<td>4.19 11.63</td>
<td>Down</td>
<td>-1.02</td>
<td>18</td>
<td>10$^6$</td>
<td>(likely) malaria resistance, thalassemia (51, 81)</td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td>rs1143627 ttgtaaagct ct</td>
<td></td>
<td></td>
<td></td>
<td>4.50 1.76</td>
<td>Up</td>
<td>0.94</td>
<td>15</td>
<td>10$^6$</td>
<td>obesity; liver, lung, gastric cancers; gastric ulcer and gastritis; Graves' disease, recurrent major depression (82–88)</td>
<td></td>
</tr>
<tr>
<td>NOS2</td>
<td>-27800550, chr17 (90) gtaaaacat at</td>
<td></td>
<td></td>
<td></td>
<td>4.50 6.78</td>
<td>Down</td>
<td>-0.42</td>
<td>8</td>
<td>10$^6$</td>
<td>(likely) rheumatoid arthritis (51, 89)</td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td>rs563763767 ccttttaga ct</td>
<td></td>
<td></td>
<td></td>
<td>1.74 1.40</td>
<td>Up</td>
<td>0.22</td>
<td>3</td>
<td>10$^2$</td>
<td>resistance to malaria, epilepsy (90, 91)</td>
<td></td>
</tr>
<tr>
<td>LEP</td>
<td>rs201381696 tcggcggcct ct</td>
<td></td>
<td></td>
<td></td>
<td>3.07 2.02</td>
<td>Up</td>
<td>0.60</td>
<td>6</td>
<td>10$^6$</td>
<td>myocardial infarction; thrombosis (92)</td>
<td></td>
</tr>
<tr>
<td>LEP</td>
<td>rs200487063 tgcctaggcc gc</td>
<td></td>
<td></td>
<td></td>
<td>3.64 12.22</td>
<td>Down</td>
<td>-1.20</td>
<td>17</td>
<td>10$^6$</td>
<td>(likely) obesity (93, 97, 96)</td>
<td></td>
</tr>
<tr>
<td>LEP</td>
<td>rs34104384 cagctataag c.t</td>
<td></td>
<td></td>
<td></td>
<td>3.64 2.39</td>
<td>Up</td>
<td>0.42</td>
<td>6</td>
<td>10$^6$</td>
<td>(likely) obesity-caused hypertension (93–95, 97)</td>
<td></td>
</tr>
<tr>
<td>ABCA9</td>
<td>rs367781716 caattttgct tc</td>
<td></td>
<td></td>
<td></td>
<td>2.90 8.11</td>
<td>Down</td>
<td>-1.02</td>
<td>15</td>
<td>10$^6$</td>
<td>(likely) higher risk of atherosclerosis, Alzheimer's disease, and cardiovascular diseases (98, 99)</td>
<td></td>
</tr>
</tbody>
</table>

$^1$Position (plus as the direct strand and minus as the complementary strand) on a chromosome within the framework of the current release GRCh38/hg38 of the reference human genome (5).

$^2$Ancestral (wt) and minor (mt) alleles.

$^3$Equilibrium dissociation constants ($K_D$) of the TATA-binding protein (TBP) binding to either the ancestral (wt) or minor (mt) allele of the human gene promoter being analyzed (nM).

$^4$SNP-caused change: “Up” and “Down” as overexpression and underexpression of the human gene considered, respectively.

$^5$Ratio of the expression levels of ancestral and minor alleles of the human gene under study, i.e., $\kappa_{\text{SNP}} = \ln(K_{\text{wt}}/K_{\text{mt}})$.

$^6$Fisher's Z-score (the ratio of the difference between two means to the square root of the sum of the squares of their standard errors).

$^7$Significance: alpha = 1 – p, where p is probability (Figure 1, B and C).

$^8$Gap, deletion.
Genomic landmarks for preventive personalized medicine

4.5. The human F3 gene

The human F3 gene encodes tissue thromboplastin (synonym: coagulation factor III). There is a known SNP marker (rs563763767) of an increased risk of venous thromboembolism and myocardial infarction in the promoter of this gene (92).

4.6. The human LEP gene

The human LEP gene encodes hormone leptin and has no known biomedical SNP markers that alter the promoter of this gene. Due to Friedman’s discovery that the LEP gene is the “obesity gene” (93), three unannotated SNPs—rs34104384, rs200487063, and rs201381696—have been predicted as candidate SNP markers of obesity-caused hypertension (94, 95) and obesity (96) as shown in Table 1 according to their predicted effects on the affinity of TBP for the LEP promoter (97).

4.7. The human ABCA9 gene

The human ABCA9 gene (ATP-binding cassette subfamily A, member 9) is one more independent example of a human gene whose promoter has no known biomedical SNP markers near its known TATA box (98) (Table 1). Using Web service SNP_TATA_Comparator (51), the candidate SNP marker rs367781716 of atherosclerosis, Alzheimer’s disease (AD), and cardiovascular diseases was both predicted in silico and experimentally verified in vitro (99).

5. PRECLINICAL VERIFICATION IN VITRO AND EX VIVO OF THE CANDIDATE SNP MARKERS OF HEREDITARY DISEASES

These predictions of Web-service SNP_TATA_Comparator (51) reviewed here were selectively verified under real-time (100–102), equilibrium (103), and nonequilibrium (97, 99, 104) conditions in vitro as well as under ex vivo conditions using cultured human cells transfected with the pGL 4.1.0 vector (Promega, USA) (reporter gene LUC (luciferase)) (105). In the text below, we discuss these experiments in more detail (for these in vitro verification experiments, full-length human TBP was expressed in Escherichia coli cells BL21 (DE3) and, then, purified by the standard protocol (106)).

5.1. Verification of candidate biomedical SNP markers using an electrophoretic mobility shift assay (EMSA) in vitro

The equilibrium dissociation constants (K_D) for the TBP–ODN complex (ODN: a given double-helical DNA 26-bp oligonucleotide synthesized and purified by BIOSYN (Novosibirsk, Russia), which is identical to either the minor or ancestral allele of a given SNP of the human gene promoter being tested, as shown in Table 1) were measured using a conventional protocol (103). It includes titration of a fixed amount of active TBP, 0.3 nM, with the increasing concentrations of the ODN to reach equilibrium. The equilibrium time was determined previously for each ODN. Binding experiments were run at 25°C in binding buffer (20 mM HEPES-KOH, pH 7.6, 5 mM MgCl_2, 70 mM KCl, 1 mM EDTA, 100 μg/ml BSA, 0.0.1% NP-40, 5% glycerol). The human TBP-ODN complexes were separated from the unbound ODN by an electrophoretic mobility shift assay (EMSA) as exemplified in Figure 3 A (ancestral allele rs367781716:T) and Figure 3 B (minor allele rs367781716:c). The K_D values that characterize the affinity of TBP to TATA boxes were determined using conventional software OriginPro 8.

Figure 3 D shows the significant correlations between the in silico predicted kappa_{SNP} values of the SNP-caused change coefficient (Table 1: natural logarithm of the ratio of quantitative estimates of the specific biological activity of ancestral (norm) and minor (mutant) alleles of the candidate SNP marker being verified) and those measured in vitro. As one can see, linear (r), rank (R and tau), generalized (gamma), and binary (p and Chi-square) correlation coefficients independently support the robustness of the predictions of Web-service SNP_TATA_Comparator (51).

Figure 4 shows the significant correlations between the in silico predicted kappa_{SNP} values of the SNP-caused change coefficient (Table 1) and those measured under the nonequilibrium conditions in vitro (97, 104). As shown in this figure, the predictions of SNP_TATA_Comparator (51) are robust.

5.2. Verification of candidate biomedical SNP markers in real time in vitro using high-resolution spectrometer SX.20

The above-mentioned ODNs were synthesized, purified, and labeled at 5'-terminals with fluorescent dyes TAMRA and FAM (BIOSYN, Novosibirsk, Russia). The working concentration of DNA duplexes in all cases was 10^{-7} M, concentration of the active TBP at the initial stage was the same, and then was increased proportionally 2-, 4-, 6-, 8-, or 10-fold. Analysis of the time-series of magnitudes was conducted next (see Figure 5 in the illustrative case of the candidate SNP marker rs201381696), and construction of a kinetic model of the interaction of TBP with each DNA duplex and calculation of velocity constants of all elementary reactions were carried out by means of the Dynafit software (Biokin, USA).

Figure 5 shows that the negative sign of the predicted kappa_{SNP} value of the SNP-caused change coefficient (Table 1) is consistent with that measured
Figure 3. Experimental verification of the candidate SNP markers (see Table 1) by an electrophoretic mobility shift assay (EMSA) under equilibrium conditions in vitro. An example of electrophorograms in the case of ancestral (panel A: norm, wild-type, wt) and minor (panel B: mutant, mt) alleles of the candidate SNP marker rs367781716 within the human ABCA9 gene promoter. (C) Experimental data: white and black squares depict the ancestral (norm, wild type, wt) and minor (mutant, mt) alleles, respectively, of the candidate SNP marker being verified as well as solid and dashed curves whose asymptotes estimate the values of the equilibrium dissociation constants (K_D) of the TBP–oligodeoxynucleotide (ODN) complexes as indicated below these curves. (D) The significant correlations between the in silico predicted kappa_{SNP} values of the SNP-caused change coefficient (Table 1: natural logarithm of the ratio of quantitative estimates of the specific biological activity of ancestral (norm) and minor (mutant) alleles of the candidate SNP marker being verified) and those measured in vitro. Solid and dashed lines or curves denote the linear regression and boundaries of its 95% confidence interval, calculated by means of software package STATISTICA (Statsoft™, USA); white and black circles depict the ancestral and minor alleles, respectively, of the candidate SNP markers being verified; r, R, tau, gamma, Chi-square, and alpha are Pearson’s simple linear correlation, Spearman’s rank correlation, Kendall’s rank correlation, the Goodman–Kruskal generalized correlation, dichotic Chi-square correlation, and their significance, respectively; p is Fisher’s exact test in the case of a 2 x 2 contingency table.

Figure 4. The significant correlations between the in silico predicted kappa_{SNP} values of the SNP-caused change coefficient (Table 1: natural logarithm of the ratio of quantitative estimates of the specific biological activity of ancestral (norm) and minor (mutant) alleles of the candidate SNP marker being verified) and those measured by EMSA under the nonequilibrium conditions in vitro. Solid and dashed lines or curves as well as r, R, tau, gamma, Chi-square, alpha, and p have the same meaning as in the legend of Figure 3.
in vitro in real time, namely: \( \kappa_{\text{SNP}} = \ln\left(\frac{4 \text{ nM}}{27 \text{ nM}}\right) = -1.91 < 0 \). This result may be one more argument in favor of the robustness of the predictions made by Web-service SNP_TATA_Comparator (51). For more details, readers can see the original articles (100–101) on other SNPs tested using high-resolution spectrometer SX.20 (Applied Photophysics, UK) and ref. (102) where another device, biosensor ProteON™ (Bio-Rad Laboratories, USA), was used.

5.3. Verification of candidate biomedical SNP markers using cultured human cell line under ex vivo conditions

Cell line HCT116 (human colon adenocarcinoma) was cultivated in a complete medium consisting of Dulbecco’s modified Eagle's medium/Nutrient mixture F-12 Ham, supplemented with 10% (v/v) of fetal bovine serum (Sigma), penicillin (100 U/mL) and streptomycin (100 mg/mL) (BioLT). The culture was maintained at 37°C in a humidified atmosphere containing 5% of CO₂ until the desired level of confluence. All the experiments were performed at 80–85% confluence. Oligonucleotides corresponding to ancestral and minor alleles of the predicted candidate SNP markers rs201381696, rs200487063, and rs34104384 (Table 1) were cloned into the pGL 4.10 vector (Promega, USA) and cotransfected with pRL-TK using Screen Fect A (InCella) as described elsewhere (107). After that, the cells were cultured in 6-well plates for 24 hours. Luciferase activity was measured by means of the Dual-Luciferase Reporter Assay kit (Promega). The results are presented in Figure 6.

As readers can see in this figure, there is a significant linear correlation (panel B) between the predicted \( \kappa_{\text{SNP}} \) values of the SNP-caused change coefficient (Table 1) and those measured in vitro in real-time mode, namely: \( \kappa_{\text{SNP}} = \ln\left(\frac{K_{D,\text{wt}}}{K_{D,\text{mt}}}\right) = \ln\left(\frac{4 \text{ nM}}{27 \text{ nM}}\right) = \ln(0.14) = -1.91 < 0 \).

Moreover, all the verification results (Figures 3–6) together indicate that this Web-service (51) is accompanied by the comprehensive pre-clinical verification platform that allows for testing its predictions as soon as they are made.

6. BIOMEDICAL PREDICTIONS USING A KEY WORDS SEARCH FOR KNOWN BIOMARKERS OF A HUMAN DISEASE IN DATABASES

The predictions in the cases of candidate SNP markers of obesity-related disorders (97), autoimmune-related pathologies (108), chronopathologies (109), aggressiveness-related comorbidities (105), and sporadic AD (49) as well as resistance to antitumor chemotherapy (110) and social dominance/submission (111) were made using Web-service SNP_TATA_Comparator (51). Below we selectively review these results in greater detail.

6.1. A Key Words search for known physiological markers of a human disease in the PubMed database

A Key Words search in the widely used PubMed database of biomedical publications (112) extends the predictive capabilities of Web service SNP_TATA_Comparator (51). Figure 7 depicts a
Genomic landmarks for preventive personalized medicine

Figure 6. Cell culture verification of candidate SNP markers rs201381696, rs200487063, and rs34104384 in human cell lines transfected with the pGL 4.10 vector carrying a reporter LUC gene. (A) Experimental data: open bars, ancestral allele (wild type, WT); gray bars, minor allele; HCT116, a human colon adenocarcinoma cell line; black bar, original vector pGL 4.10 (Promega, USA) without any insertions as an independent control. The height of the gray bars and their error bars correspond to the mean estimates and boundaries of their 95% confidence intervals calculated from at least three independent measurements. Asterisks indicate a statistically significant difference at the confidence level alpha < 0.05. (B) Significant correlation between the \textit{in silico} predicted \(\kappa_{\text{SNP}}\) values of the SNP-caused change coefficient (Table 1: natural logarithm of the ratio of quantitative estimates of the specific biological activity of ancestral (white circles, norm, wt) and minor (black circles, mutant, mt) alleles of the candidate SNP marker being verified) and those measured \(\text{ex vivo}\). Solid and dashed lines or curves denote the linear regression and boundaries of its 95% confidence interval, calculated by means of software package STATISTICA (Statsoft\textsuperscript{TM}, USA); \(r\) and alpha are Pearson’s simple linear correlation coefficient and its statistical significance level, respectively.

6.2. Examples of SNP markers of hereditary diseases as candidate SNP markers of sporadic Alzheimer’s disease (AD)

Table 2 exemplifies the results of the Key Words search for sporadic AD (49, 113–148) as shown in Figure 7. Below we discuss only some of these results in detail. The candidate SNP marker (rs72661131) of sporadic AD as a consequence of stroke (49) is quite interesting in our opinion. It is a genomewide marker suggesting that people who carry the minor allele of this SNP could minimize the risks of both diseases by adding some natural marine products into their diet (50). Of course, this is oversimplification: if stroke and sporadic AD could be prevented by a natural marine diet alone, the life of the general population would be very easy. Similarly, using the candidate SNP marker rs567653539, someone with a minor allele of this SNP can prevent cardiovascular disease, rheumatoid arthritis, and sporadic AD by switching to a diet enriched with selenium (Se) (119). In addition, candidate SNP markers (rs33931746, rs63750953, rs281864525, rs35518301, and rs34166473) of sporadic AD as a comorbidity of thalassemia may mean that the functional food supplement fermented papaya can prevent these diseases (125).

Thus, these genomewide landmarks may be interesting for people who would like to bring their lifestyle in line with their sequenced individual genome.
Genomic landmarks for preventive personalized medicine

Table 2. Examples of SNP markers of hereditary diseases that can be candidate SNP markers of sporadic Alzheimer’s disease (AD)

<table>
<thead>
<tr>
<th>Genes</th>
<th>dbSNP (6)</th>
<th>hereditary diseases whose known and candidate (indicated by “(likely)”) SNP markers are being analyzed</th>
<th>predicted obesity-related comorbidities whose candidate SNP markers (Figure 7: Result-2) are indicated by “(likely)”</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBL2</td>
<td>rs72661131, rs562962093</td>
<td>variable immunodeficiency, preeclampsia, stroke, and (likely) cardiovascular risk in rheumatoid arthritis</td>
<td>(likely) higher risk of sporadic AD, human intravenous immunoglobulin is a treatment of variable immunodeficiency that may account for its beneficial effect in both familial and sporadic AD; both preeclampsia and sporadic AD are mapped onto the human genome region 10q22 where the preeclampsia-associated STOX1 gene (isoform A) promotes the growth of beta-amyloid plaques; natural marine diet prevents both stroke and sporadic AD</td>
<td>(49, 50, 113–115)</td>
</tr>
<tr>
<td>MBL2</td>
<td>rs567653539</td>
<td>(likely) cardiovascular risk in rheumatoid arthritis</td>
<td>(likely) lower risk of sporadic AD whereas both cardiovascular diseases and rheumatoid arthritis positively correlate with AD; circulating microRNAs as a biomarker of AD, cardiovascular diseases, and rheumatoid arthritis; selenium (Se) can prevent and/or treat a cardiovascular disease, rheumatoid arthritis, and AD.</td>
<td>(49, 113, 116–119)</td>
</tr>
<tr>
<td>HBB, HBD</td>
<td>rs33980857, rs34598520, rs33931746, rs63750953, rs281864525, rs35518301, rs34166473</td>
<td>malaria resistance, thalassemia</td>
<td>(likely) in homozygotes, high risk of AD (both under- and over-expression of either HBB or HBD genes are associated with an increased risk of AD and more rapid cognitive decline (i.e., abnormal hemoglobin abundance correlates with both AD and cognitive decline), whereas beta-amyloid and hemoglobin aggregate with each other near traumatic vascular injury in the brain); heterozygote protects from AD; the functional food supplement fermented papayas as a treatment of both thalassemia and sporadic AD; drugs reducing Fe concentration can treat both thalassemia and sporadic AD; manzamine-type alkaloids as drugs effective against both malaria and sporadic AD; some alleles of APOE (keynote gene of both familial and sporadic AD pathogenesis) prevent malaria in race-biased susceptibility to AD</td>
<td>(49, 120–128)</td>
</tr>
<tr>
<td>IL1B</td>
<td>rs1143627</td>
<td>obesity; liver, lung, gastric cancers; gastric ulcer and gastritis; Graves’ disease, recurrent major depression</td>
<td>(likely) greater beta-amyloid plaque clearance and blood-brain barrier damage in AD; sporadic AD positively correlates with liver and gastric cancers rather than lung cancer; negative correlation between sporadic AD and Graves’ disease; same antidepressant drugs are used during recurrent major depression and both familial and sporadic AD; positive correlation between sporadic AD and obesity, whereas both body weight and fat loss regardless of diets and lifestyle are biomarkers of progression of both familial and sporadic AD</td>
<td>(49, 129–141)</td>
</tr>
<tr>
<td>IL1B</td>
<td>rs549858786</td>
<td>(likely) rheumatoid arthritis</td>
<td>(likely) lower beta-amyloid plaque clearance and blood-brain barrier damage in AD</td>
<td>(49, 129, 130)</td>
</tr>
<tr>
<td>F3</td>
<td>rs563763767</td>
<td>myocardial infarction; thrombosis</td>
<td>(likely) high risk of AD because F3 is found in β-amyloid plaques; donepezil is a drug against both sporadic AD and myocardial infarction; pathogenesis of sporadic AD and myocardial infarction share similarities; fibrinogen and beta-amyloid bind to each other as a risk factor for both thrombosis and both familial and sporadic AD; drug RU-505 significantly alleviates thrombosis and cognitive deficits in familial and sporadic AD</td>
<td>(49, 142–146)</td>
</tr>
<tr>
<td>LEP</td>
<td>rs201381696</td>
<td>(likely) obesity</td>
<td>(likely) positive correlation between sporadic AD and obesity, whereas both body weight and fat loss regardless of diets and lifestyle are biomarkers of progression of AD; higher risk of AD where beta-amyloid aggregates can cause hypothalamic leptin signaling dysfunction leading to early body weight loss; AD treatment involves nutritional assessments and dietary measures; there is leptin replacement therapy for AD in case of leptin deficiency and weight loss</td>
<td>(49, 93, 140, 141, 147, 148)</td>
</tr>
<tr>
<td>LEP</td>
<td>rs200487063, rs34104384</td>
<td>(likely) obesity-caused hypertension</td>
<td>(likely) significant positive correlation between sporadic AD and obesity, whereas both body weight and fat loss regardless of diets and lifestyle are biomarkers of progression of both familial and sporadic AD; hyperleptinemia in urban children can elevate risk of sporadic AD in the elderly</td>
<td>(49, 93, 147)</td>
</tr>
</tbody>
</table>

6.3. Examples of SNP markers of hereditary diseases as candidate SNP markers of autoimmunity-related comorbidities

Table 3 demonstrates the results of the Key Words search for autoimmunity-related comorbidities (108, 149–169). In our opinion, the candidate SNP marker (rs563763767) of Hughes syndrome-associated thrombosis (lethal during pregnancy (167, 168)) is the most interesting because this syndrome is easy to diagnose early and is a preventable form of myocardial ischemia (169). Using this genomewide
landmark, a pregnant woman carrying the minor allele of this SNP and her physician can arrange additional diagnostics to detect the onset and development of this disease during her pregnancy to prevent its fatal outcome.

6.4. Examples of SNP markers of hereditary diseases as candidate SNP markers of circadian rhythm-related comorbidities

Table 4 contains the results of the Key Words search for circadian rhythm-related comorbidities (170–182), which are associated with consequences of desynchronoses either among the nervous, immune, digestive, respiratory, and other systems of the human body or between the human body and its environment (109). The most interesting (in our opinion) known SNP marker of myocardial infarction and thrombosis (rs563763767) may indicate that these hereditary diseases have a circadian preference for the early morning (182), which should correspond to the period of the strongest therapeutic effects of the medication used. In addition, candidate SNP markers rs33980857, rs34598529, rs33931746, rs63750953, rs21864525, rs3518301, and rs34166473 may indicate that during deferoxamine-based therapy in thalassemia, a patient can develop symptoms of sensorineural hearing loss as a complication of this drug (170, 171). Therefore, additional analysis of hearing capacity in patients during this treatment may prevent this hearing loss (170, 171).

6.5. Examples of SNP markers of hereditary diseases as candidate SNP markers of obesity-related comorbidities

Table 5 shows the results of the above-mentioned Key Words search on candidate SNP markers of obesity-related disorders (82, 94–97, 183–188). Looking through this table, one can see that obesity is a common comorbidity of all the hereditary diseases that have been selected for this review. The most generally accepted point of view is that obesity can lead to complications of many other diseases. Thus, our results suggest that prevention of obesity may further alleviate many other diseases, in line with the point of view above.

6.6. Examples of SNP markers of hereditary diseases as candidate SNP markers of aggressiveness-related complications

Table 6 exemplifies the results of the Key Words search for aggressiveness-related complications (105, 189–210), which are associated with a hereditary behavioral trait that mobilizes all systems of the body—i.e., nervous, endocrine, respiratory, vascular, and muscular systems—for the defense of oneself, children, family, home, territory, possessions, business, ideas, or interests (211). It is noteworthy that aggressiveness is a specific powerful internal stressor (whereby the human body mobilizes itself) that can negatively affect the lifespan (212). Here, it should be
Table 4. Examples of SNP markers of hereditary diseases that can be candidate SNP markers of circadian rhythm-related comorbidities

<table>
<thead>
<tr>
<th>Genes</th>
<th>dbSNP (6) or GRCh38/ hg38† (Reference)</th>
<th>hereditary diseases whose known and candidate (indicated by “(likely)”) SNP markers are being analyzed</th>
<th>predicted obesity-related comorbidities whose candidate SNP markers (Figure 7: Result-2) are indicated by “(likely)”</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBB,</td>
<td>rs33990857, rs34598529, rs33931746, rs63750953, rs281864525, rs35518301, rs34166473</td>
<td>malaria resistance, thalassemia</td>
<td>(likely) circadian symptoms (worse at night) in restless legs syndrome caused by iron deficiency anemia co-occurring with thalassemia and for sensorineural hearing loss as a complication of deferoxamine-based therapy in thalassemia</td>
<td>(109, 170–174)</td>
</tr>
<tr>
<td>HBD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td>rs1143627</td>
<td>liver, lung, gastric cancers; gastritis; gastric ulcer; Graves’ disease, recurrent major depression</td>
<td>both diagnosis and treatment of these hereditary diseases have circadian optima for use; (likely) bipolar disorder whose diagnosis and treatment have circadian optima depending on the diet</td>
<td>(109, 175–177)</td>
</tr>
<tr>
<td>IL1B</td>
<td>rs549858796</td>
<td>(likely) rheumatoid arthritis</td>
<td>(likely) this hereditary disease disrupts the circadian rhythm of IL1B gene expression</td>
<td>(109, 178, 179)</td>
</tr>
<tr>
<td>NOS2</td>
<td>-27800550, chr17 (90)</td>
<td>resistance to malaria, epilepsy</td>
<td>epilepsy that damages the hypothalamus and the circadian clock as a whole (clinically known) (likely) remission of panic disorder whose symptoms are circadian (worse late in the evening)</td>
<td>(90, 109, 180, 181)</td>
</tr>
<tr>
<td>F3</td>
<td>rs563763767</td>
<td>myocardial infarction; thrombosis</td>
<td>these hereditary diseases are characterized by their circadian preference for the early morning in the elderly</td>
<td>(109, 182)</td>
</tr>
</tbody>
</table>

Table 5. Examples of SNP markers of hereditary diseases that can be candidate SNP markers of obesity-related comorbidities

<table>
<thead>
<tr>
<th>Genes</th>
<th>dbSNP (6) or GRCh38/ hg38† (Reference)</th>
<th>hereditary diseases whose known and candidate (indicated by “(likely)”) SNP markers are being analyzed</th>
<th>obesity-related comorbidities whose candidate SNP markers (Figure 7: Result-2) are indicated by “(likely)”</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBL2</td>
<td>rs72661131, rs562962093</td>
<td>variable immunodeficiency, stroke, preeclampsia, and (likely) cardiovascular risk in rheumatoid arthritis</td>
<td>(likely) obesity</td>
<td>(97, 183)</td>
</tr>
<tr>
<td>MBL2</td>
<td>rs56765359</td>
<td>(likely) cardiovascular risk in rheumatoid arthritis</td>
<td>(likely) obesity</td>
<td>(97, 183)</td>
</tr>
<tr>
<td>HBB,</td>
<td>rs33990857, rs34598529, rs33931746, rs63750953, rs281864525, rs35518301, rs34166473</td>
<td>malaria resistance, thalassemia</td>
<td>(likely) obesity-caused chronic inflammation</td>
<td>(97, 184)</td>
</tr>
<tr>
<td>HBD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td>rs1143627</td>
<td>liver, lung, gastric cancers; gastric ulcer and gastritis; Graves’ disease, major recurrent disorder</td>
<td>obesity (clinically known)</td>
<td>(51, 82)</td>
</tr>
<tr>
<td>NOS2</td>
<td>-27800550, chr17 (90)</td>
<td>resistance to malaria, epilepsy</td>
<td>(likely) obesity</td>
<td>(90, 97, 185)</td>
</tr>
<tr>
<td>F3</td>
<td>rs563763767</td>
<td>myocardial infarction; thrombosis</td>
<td>(likely) leptin replacement therapy can prevent weight gain and obesity</td>
<td>(97, 186)</td>
</tr>
<tr>
<td>LEP</td>
<td>rs201381696</td>
<td>(likely) obesity</td>
<td>(likely) hyperleptinemia promotes the development of hypertension during pregnancy (rat model)</td>
<td>(93, 96, 187)</td>
</tr>
<tr>
<td>LEP</td>
<td>rs200487063, rs34104384</td>
<td>(likely) obesity-caused hypertension</td>
<td></td>
<td>(93–95, 97, 188)</td>
</tr>
</tbody>
</table>

Position (plus as the direct strand and minus as the complementary strand) on a chromosome within the framework of the current release GRCh38/ hg38 of the reference human genome (5).

emphasized that known SNP markers of hereditary diseases are the cause of the diseases in question. In contrast, candidate SNP markers of polygenic traits (e.g., aggressiveness) can only serve as genomewide informative markers of the increased or decreased risk relative to the norm among people who carry the minor alleles of these SNPs (108). For example, using one whole-genome marker (rs201381696) of childhood-aggressiveness, parents of an aggressive 10-year-old girl with a minor allele of this candidate SNP marker may choose a diet and a physical exercise regimen for their daughter to prevent her obesity in adolescence and cardiovascular complications in adulthood (202). Similarly, using a candidate SNP marker (rs1143627) of aggressiveness as a complication of cytokine immunotherapy, a physician can prescribe an adjuvant
antiaggression medication together with this main treatment procedure to a patient carrying a minor allele of this SNP (189, 196). Candidate SNP markers (rs33980857, rs34598529, rs33931746, rs63750953, rs281864525, rs35518301, and rs34166473) of aggressiveness as a complication of hemodialysis in severe thalassemia can be used in a similar fashion (189, 196). Furthermore, regarding the candidate SNP marker of higher aggressiveness in males subjected to environmental pollution with Pb (the –51T>C substitution, NOS2 gene promoter), people with a minor allele of this SNP can alter their lifestyle to minimize their contact with materials containing lead (197). Therefore, this information could be interesting to people who would like to bring their lifestyle in line with their sequenced individual genome.

7. SUMMARY AND PERSPECTIVE

Currently, the widely used public database PubMed contains a huge number of clinical cases, retrospective reviews, research articles, laboratory data, and empirical findings from clinicians, nutritionists, pharmacists, physiologists, geneticists, psychologists, bioinformaticians, pedagogues, sociologists, legal scholars, economists, and other relevant experts such as specialists on insurance, management, health care, law enforcement, and

Table 6. Examples of SNP markers of hereditary diseases that can be candidate SNP markers of aggressiveness-related comorbidities

<table>
<thead>
<tr>
<th>Genes</th>
<th>dbSNP (6) or GRCh38/hg38 (Reference)</th>
<th>hereditary diseases whose known and candidate (indicated by “(likely)” SNP markers are being analyzed</th>
<th>aggressiveness-related comorbidities whose candidate SNP markers (Figure 7: Result-2) are indicated by “(likely)”</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBB, HBD</td>
<td>rs33980857, rs34598529, rs33931746, rs63750953, rs281864525, rs35518301, rs34166473</td>
<td>malaria resistance, thalassemia</td>
<td>(likely) thalasemia-related male–male aggression, socialized aggression, inattention, low IQ, acute psychosis with aggression, impulsiveness as a form of aggressiveness; aggression as a comorbidity in both 4-yo girls and 5-yo boys hospitalized with thalassemia; aggressiveness as a consequence of hemodialysis in severe thalassemia</td>
<td>(105, 189–195)</td>
</tr>
<tr>
<td>IL1B</td>
<td>rs1143627</td>
<td>liver, lung, gastric cancers; gastric ulcer and gastritis; Graves’ disease, recurrent major depression</td>
<td>(likely) highly aggressive traits in patients who receive cytokine immunotherapy or regular hemodialysis (clinical retrospective review)</td>
<td>(105, 189), 196</td>
</tr>
<tr>
<td>IL1B</td>
<td>rs549858786</td>
<td>(likely) rheumatoid arthritis</td>
<td>(likely) less aggressive traits in patients who receive cytokine immunotherapy or regular hemodialysis (clinical retrospective review)</td>
<td>(105, 189), 196</td>
</tr>
<tr>
<td>NOS2</td>
<td>-27800550, chr17 (90)</td>
<td>resistance to malaria, epilepsy</td>
<td>(likely) aggressiveness as a complication of both drug-resistant and childhood epilepsy; stigma as a critical factor for interictal aggression in epilepsy (clinical review); aggression, hyperactivity, and impaired memory coexist in case of recurrent spontaneous seizures in epilepsy; a gender-biased complication of excessive lead (Pb) intake manifested as lesser exploration in females and higher aggressiveness in males</td>
<td>(90, 105, 197–201)</td>
</tr>
<tr>
<td>LEP</td>
<td>rs201381696</td>
<td>(likely) obesity</td>
<td>(likely) in a 10-yo girl, aggressiveness is a predictive factor for prevention of obesity in adolescence with cardiovascular complications in adulthood, as is the case for 5-yo boys (retrospective review); aberrant maternal behavior, low aggression against an unknown social stimulus and locomotor activity during a high-fat diet; low probability of dominance due to aggressiveness against subordinates in females; high risks of suicidality, violence, and impulsive aggressiveness in schizophrenia; higher social aggressiveness in males; longer survival in aggressive leptin-deficient women with anorexia nervosa</td>
<td>(93, 105, 202–210)</td>
</tr>
<tr>
<td>LEP</td>
<td>rs200487063, rs34104384</td>
<td>(likely) obesity-caused hypertension</td>
<td>(likely) lower risk of aberrant maternal behavior; higher aggression against an unknown social stimulus, and locomotor activity on a high-fat diet; higher probability of dominance due to aggressiveness against subordinates in females; lower risks of suicidality, violence, and impulsive aggressiveness in schizophrenia; lower social aggressiveness in males</td>
<td>(93, 105, 203–210)</td>
</tr>
</tbody>
</table>

1Position (plus as the direct strand and minus as the complementary strand) on a chromosome within the framework of the current release GRCh38/hg38 of the reference human genome (5).
environmental protection. The gigantic scale, multidisciplinary nature, complexity, and disarray of this information pool may hinder the practical use of this vital knowledge by the general population. As one can see in Tables 1–6, candidate biomedical SNP markers seem to be promising whole-genome landmarks. Using these markers, researchers can subdivide existing knowledge on the relevant biomedical SNP markers into readable portions, which may be directly applicable to people carrying a minor allele of such SNPs. This approach could be interesting for people who would like to bring their lifestyle in line with their sequenced individual genome.

8. ACKNOWLEDGEMENTS

We are grateful to Shevchuk Editing (Brooklyn, NY, USA; http://www.shevchuk-editing.com) for editing of the English text. Manuscript writing was supported by grant #14.B25.3.1.0.033 from the government of the Russian Federation, Resolution No. 220 (for M.P.). The software development was supported by project #14-24-00123 from the Russian Sciences Foundation (for D.R. and N.K.). The data compilation was supported by project #17-04-00501 from the Russian Foundation for Basic Research (for I.C. and E.K.). The data analysis was supported by project #17-04-00505 from the Russian Foundation for Basic Research (for E.S.). The interpretation of the results was supported by project #17-44-543344 from the Russian Foundation for Basic Research (for I.C.). The concept and study design were supported by project #0324-2015-0003 of the Russian Government Budget (for L.S., N.I., D.Z., and O.A.).

9. REFERENCES

Genomic landmarks for preventive personalized medicine


Genomic landmarks for preventive personalized medicine


Genomic landmarks for preventive personalized medicine


51. Ponomarenko, M., Rasskazov, D., Arkova, O., Ponomarenko, P., Suslov, V., Savinkova, L., and Kolchanov N.: How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene’s promoter via a change in affinity of the TATA-binding
DOI: 10.1155/2015/359835

DOI: 10.1101/gr.361602

DOI: 10.1074/jbc.270.23.13850

DOI: 10.1093/bioinformatics/12.5.441

DOI: 10.1093/bioinformatics/15.7.687

DOI: 10.1016/0022-2836(87)90354-8

DOI: 10.1016/0022-2836(90)90223-9

DOI: 10.1016/S0006-3495(98)77521-6

DOI: 10.1074/jbc.M110147200


DOI: 10.1073/pnas.86.15.5718

DOI: 10.1093/bioinformatics/15.7.687


DOI: 10.1016/0022-2836(71)90319-6

DOI: 10.1142/S0219720010004677

DOI: 10.1186/s12859-015-0701-2

67. Ponomarenko, P.M., Savinkova, L.K., Drachkova, I.A., Lysova, M.V., Arshinova,
Genomic landmarks for preventive personalized medicine


78. Troelsen, L.N., Garred, P., Christiansen, B., Torp-Pedersen, C., Christensen, I.J., Narvestad, E., and Jacobsen, S.: Double role of mannose-binding lectin in relation to carotid intima-media thickness in patients
DOI: 10.1016/j.molimm.2009.10.021


DOI: 10.1038/oby.2007.95

DOI: 10.1001/jhep.2003.50017

DOI: 10.1016/j.cca.2010.05.035

DOI: 10.1038/35006081

DOI: 10.1186/1471-230X-10-126

DOI: 10.1111/j.1600-0528.2009.01538.x

DOI: 10.1007/s12031-011-9507-5

DOI: 10.1007/s00296-011-1825-y

DOI: 10.1038/sj.gene.6364022

DOI: 10.3171/2008.6.17608

DOI: 10.1161/01.ATV.20.3.892

DOI: 10.1038/372425a0

DOI: 10.1097/HJH.0b013e3282e9a9fd

DOI: 10.1097/HJH.0b013e3282f47688

DOI: 10.1007/s00401-013-1190-x

DOI: 10.1186/1471-2164-16-S13-S5

DOI: 10.1093/nar/gku1111

DOI: 10.1134/S2079059716070029


101. Arkova, O., Kuznetsov, N., Fedorova, O., and Savinkova, L.: A real-time study of the interaction of TBP with a TATA box-containing duplex identical to an ancestral or minor allele of human gene LEP or TPI. *J Biomol Struct Dyn* 2016;
DOI: 10.1.080/07391102.2.016.1.241190


DOI: 10.1371/journal.pone.0054626


DOI: 10.1186/s12864-016-3353-3

DOI: 10.1385/0-89603-288-4:359

Genomic landmarks for preventive personalized medicine

DOI: 10.1210/mend.16.3.0780

DOI: 10.3389/fimmu.2016.00130

DOI: 10.1155/2016/8642703

DOI: 10.1134/S0026893316010209

DOI: 10.18699/VJ16.196

DOI: 10.1093/nar/gku1130


DOI: 10.1186/1742-2094-9-105


DOI: 10.1155/2015/471342


DOI: 10.3389/fgene.2013.00121

DOI: 10.2174/092986711797379249


DOI: 10.1212/WNL.0b013e318225a9a9


137. Akushevich, I., Kravchenko, J., Ukrainsteva, S., Arbeev, K., Kulminski, A., and Yashin,


151. Filipovic, B., Sporicic, Z., Randjelovic, T., and Nikolic, G.: Common variable immunodeficiency associated with
Genomic landmarks for preventive personalized medicine


167. Sun, Y., Cui, M., Zhu, W., Xu, W., and Li, N.: A case report of a pregnancy-related death caused by primary antiphospholipid...
DOI: 10.2147/IMCRJ.S71321

DOI: 10.1177/107602960601200107

DOI: 10.1016/j.neuint.2006.11.009

DOI: 10.1007/s00296-011-1825-y

DOI: 10.1111/j.1742-1241.2010.02340.x


DOI: 10.1159/000276227


DOI: 10.1155/2014/293216

DOI: 10.1017/S0022215107000874

DOI: 10.1186/1471-244X-13-77

DOI: 10.1210/jc.2014-3868

DOI: 10.1016/j.smrv.2009.01.002

DOI: 10.1080/07420500701282307


**Abbreviations:** ABCA9: ATP-binding cassette subfamily a, member 9; AD: Alzheimer’s disease; EMSA: electrophoretic mobility shift assay; F3: coagulation factor III (synonyms: thromboplastin, tissue factor); GRCh38/hg38: the current release of the reference human genome; HBB: hemoglobin subunit beta; HBD: hemoglobin subunit delta; IL1B: interleukin 1 beta (synonyms: catabolin); LEP: leptin (synonyms: satiety hormone, obesity factor (OB)); MBL2: mannose-binding lectin 2 (synonyms: protein C, collectin-1); NOS2: nitric oxide synthase 2 (inducible); PPPM: predictive preventive personalized medicine; SNP: single nucleotide polymorphism; TATA box: TBP-binding site; TBP: TATA-binding protein; TSS: transcription start site.

**Key Words** Predictive Preventive Personalized Medicine, Gene, Promoter, TATA-binding protein, TATA box, single nucleotide polymorphism, SNP, Overexpression, Deficiency, Hereditary Disease, Comorbidity, Complication, Biomedical SNP marker, Review

**Send correspondence to:** Mikhail Ponomarenko, Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences, 10 Lavrentyev Avenue, Novosibirsk, 630090, Russia. Tel: 7-383-3634923, Fax: 7-383-3331278, E-mail: pon@bionet.nsc.ru